Temporal Data and Data Temporality: Time is change, not only order.

Catherine Letondal
letondal@lri.fr

INRIA Saclay
F-91405, Orsay, France

Aurélien Tabard
aurelien@tabard.fr

LRI - Univ. Paris-Sud & CNRS
F-91405, Orsay, France

Wendy Mackay
mackay@lri.fr

TELECOM ParisTech – CNRS LTCI
F-75013, Paris, France

Abstract
In this paper we describe two approaches to temporal data and how taking both into account can help improve the design of temporal data-based tools. *Time as order* considers temporal data as data that can be described by a time attribute, which can improve navigation or organization. *Time as change* considers temporal data as data that evolve over time.

Keywords
Time, temporal data, reflection, laboratory notebooks.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction

The majority of tools supporting temporal data use time as a convenient property to help search data, or to get a global view on a series of information items. As such, time helps supporting navigation and organization tasks. In this paper, we would like to emphasize a different aspect, namely the fact that data itself evolve over time. This corresponds to a classical and quite well-established distinction in the history of ideas: time as order, and time as change [Wolff 2004]. The former is usually the physicists' view of time, as an objective

non-reversible arrow, where time is the fact that everything happen in a global ordering. The latter belongs to a more philosophical approach, where the focus is on the experience of time, mostly through the observation that changes, such as aging or processing, occur.

Using time attributes associated to data for navigation or organization purposes rely on the time as order paradigm. Focusing on processes and data changes takes the time as change approach. Temporal data can thus be described either as data having time attributes, or data needing a dynamic description, acknowledging data temporal dimension - what we call temporality.

Data temporality in the context of biology research

The work we present is grounded in our observation of biologists' relationship to information. We focus here on the way they organize information through their laboratory notebooks and various other tools, move it from one medium to another, and revisit it. We illustrate how the *time as change* paradigm helps to better capture the dynamic aspects of their data through three types of data temporality: TODO lists, reflective activity (biology research) and project management.

TODO items are a familiar type of temporal and dynamic data. Indeed, a TODO is an item that follows a typical path: to be done, with or without a deadline, then maybe imminent when hitting the deadline, then done. After being done, a TODO item can then be either thrown, or archived.

Scientific ideas also follow a path from their beginning through various steps that may include blackboard diagrams, discussions and meetings notes, experimental data, peer scientific publications, etc… This is related to the fact that biology researchers are in a creative process and **reflect** on their decisions in order to explore new leads or justify their decisions. Paper laboratory notebooks show this temporality of thoughts.

At a higher level, with their roadmaps and deadlines, **projects** also hold a temporal dimension. Yet projects and time organizations are often in conflict. Both provide dimensions to discuss activity but are they difficult to integrate [Tabard & al, 2007]. As researchers explore different projects, it is difficult to put order in their ideas, their framing changes and transitions or regrouping occur [figure 1] (*time as change*). When the projects are over, defined, researchers can refer to them as a whole and situate them in time (*time as order*).

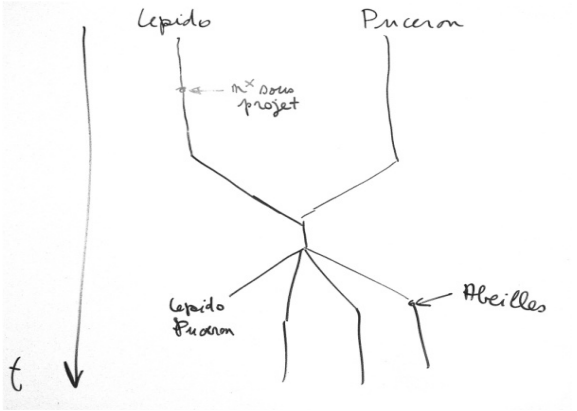

Figure 1: evolution of genomic projects organization over time. Common analysis to both projects are joined in one project, whereas other activities are spiltted.

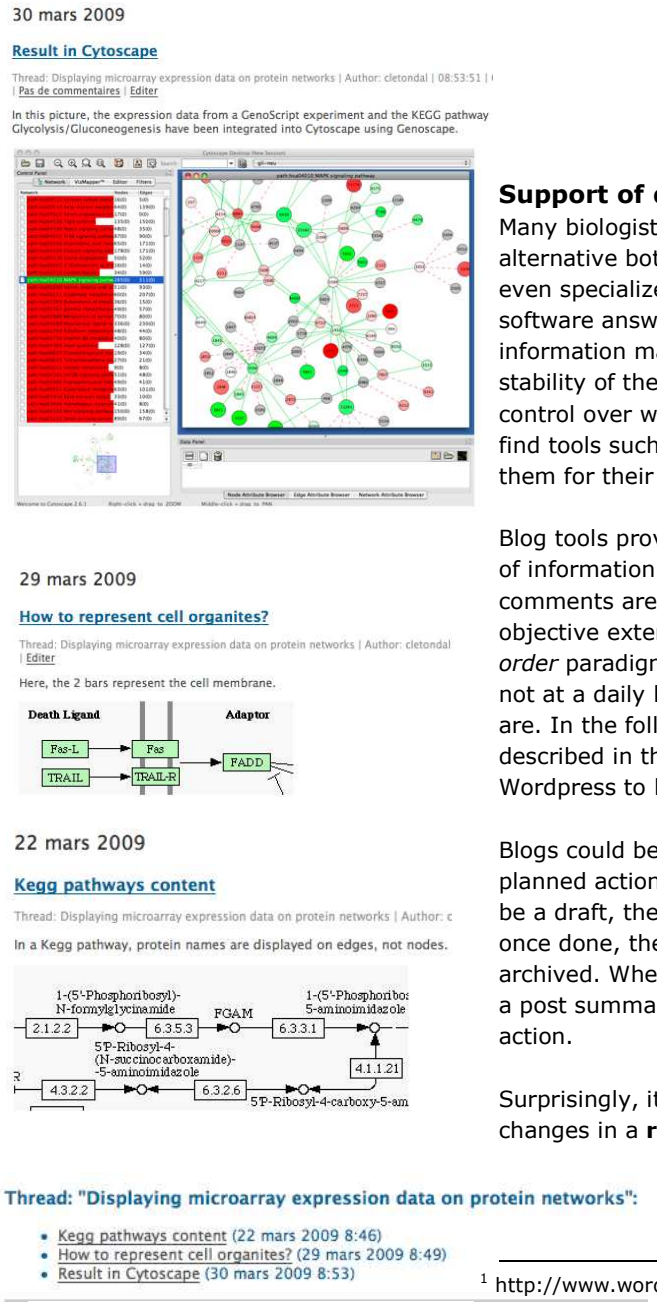

Figure 2: Thread of the evolution of a bioinformatics experiment [Clément-Ziza & al. 2009]

Support of data temporality in blogs

Many biologists are exploring blogging tools as an alternative both to paper laboratory notebooks and even specialized electronic notebooks. Blogging software answer three of their concerns regarding information management: durability of the data, stability of the system, confidentiality and privacy control over what they wrote. Furthermore, biologists find tools such as Wordpress[1] flexible enough to adapt them for their own need.

Blog tools provide a strong support for the chronology of information. Each post together with edits and comments are time-stamped, date becomes an objective external attribute. Blogs thus fit the *time as order* paradigm. Yet, they are organized at a post level, not at a daily level by default like laboratory notebooks are. In the following, we revisit the 3 types of data described in the previous part and see how we adapted Wordpress to better suit laboratory work.

Blogs could better support and integrate ongoing and planned actions. In the blog model, a **TODO item** can be a draft, then an email if a deadline is met. Then, once done, the item can either be thrown away or archived. When appropriate, the researcher may create a post summarizing the outcome of the corresponding action.

Surprisingly, it is quite difficult to express evolution and changes in a **reflective activity** over several posts: follow-ups of a post are available only through comments or trackbacks, which are second order writing - they do not appear as proper items in the blog index. Wordpress provides versioning, which enables to manage items that need to be changed. However, updating a post and just having access to its changes history does not capture the process over time, within the global stream of posts. Blog tools are designed as publishing tools; they do not support iterative thinking the way paper notebooks do.

Threading seems more appropriate to express data changes and processes. For example the evolution from raw data, to filtered data, to preliminary analysis of sample, to statistical analysis of large scale experiment. Everything refers to the same experiment, the data is the same, yet the information is changing. A biologist reporting an experiment lasting over several days or even weeks can keep track of the same "item" - the experiment - through following posts. Figure 2 shows a thread of 3 posts related to the same bioinformatics experiment. Threads are different from the idea of projects or categories. Threads are not defined in advance it is the information with the thread that define them. On the contrary **projects** or categories define the content they hold.

Medium temporality

Contrary to paper notes, computer files do not display the traces or versions that led to their final state. The transparent iterative edition capabilities of computer files hide their history to the users. Because of computers' editing flexibility, the transitions from action items, to reflection, to finished and articulated project, are lost.

It is thus important to remember that the medium and the information interact. The medium influence what

[1] http://www.wordpress.org

biologist will save, while biologists pick the medium that fits best their needs. Paper scraps suit short notes, whereas biologists use computer tools to document their digital activity.

Paper holds temporal properties which are not yet integrated in computer. Paper notes and notebooks display their age, the number of iterations they went through. Furthermore, users can program paper decay: rather than pasting a sheet in a notebook, leaving it loose hints that it will disappear at some point in time.

As more activity happens online, biologists use less paper notebooks. However they feel they "lose discipline" when moving to complete digital notebooks. Paper allows smooth transitions between action items to reflective activity and project management. When saving a short note that may have no value later, paper scraps are still more efficient. Paper notebooks support easy re-visitation and reflection, by providing a constrained writing, with a limited space (only shorts edits are possible) and chronological order.

Compared to paper artifacts such as laboratory notebooks, computer files do not offer a proper structure to manage temporal and evolutive data. Being editable, it is difficult to reach a definitive state of a computer-based document. This is one of the reasons why biologists still use paper notebook for their linear and constraining structure.

Computer-based notebooks should provide a similar temporal structure, as blogging tools do with their automatic time-stamping of posts. Even in the presence of discordance between an automatic timestamp and the actual date of an experiment, we observed that biologists prefer to keep this automatic and objective date, rather than modify it - and manually add the *real* date of the experiment.

Conclusion

Providing tools to navigate and search among time-stamped data is not enough. There is also a need for mechanisms to support transformations and processes over time, both for scientific data and scientific ideas. These mechanisms should not only help the user visualize but also *express* time and change.

Citations

[1] Clément-Ziza, C., Malabat, C., Weber, C., Moszer, I., Aittokallio, T., Letondal, C. and Rousseau, S., 2009 Genoscape: a Cytoscape plug-in to automate the retrieval and integration of gene expression data and molecular networks. Bioinformatics, Oxford University Press, submitted

[2] Wolff, Francis. Dire le monde, Paris, P.U.F., 1997, 220 p. 2ème édition complétée avec une Préface inédite et deux annexes, P.U.F., « Quadrige », 2004, 284 p.

[3] Tabard, A., Mackay, W. E., and Eastmond, E. 2008. From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. In *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work.* CSCW '08. ACM, 569-578.